Document generated on 07/04/2025 1:28 p.m.

Documentation et bibliothèques

Télédéchargement et création de bases de données locales : communication, conversion et droit d'auteur

Downloading and data base creation: communication, conversion and copyright

Teledescarga y creación de bases de datos: comunicación, conversión y derechos de autor

Yvette Henry-Rousseau, Fethy Mili and Sylvie Cournoyer

Volume 32, Number 4, October–December 1986

URI: https://id.erudit.org/iderudit/1052615ar
DOI: https://doi.org/10.7202/1052615ar

See table of contents

Publisher(s)

Association pour l'avancement des sciences et des techniques de la documentation (ASTED)

ISSN

0315-2340 (print)
2291-8949 (digital)

Explore this journal

Cite this article

Henry-Rousseau, Y., Mili, F. & Cournoyer, S. (1986). Télédéchargement et création de bases de données locales : communication, conversion et droit d'auteur. *Documentation et bibliothèques*, *32*(4), 117–134. https://doi.org/10.7202/1052615ar

Article abstract

The downloading of data for the purposes of creating local data bases is of growing interest to information professionals. The following article examines this phenomenon.

Following a historical summary, the first part of the article describes the creation of a data base after downloading. All the steps are outlined and a number of telecommunications software are described.

The second part defines and describes the conversion. An emphasis is placed on the technical aspects involved in the processing of the downloaded data. Several examples illustrate this process.

The third part briefly summarises the American Copyright law and the legal problems arising from different types of downloading. The concerns of the producers and vendors are outlined as well as the means taken to control downloading.

Tous droits réservés © Association pour l'avancement des sciences et des techniques de la documentation (ASTED), 1986

This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

https://apropos.erudit.org/en/users/policy-on-use/

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

https://www.erudit.org/en/

Télédéchargement et création de bases de données locales: communication, conversion et droit d'auteur

Yvette Henry-Rousseau*
Fethy Mili*
Sylvie Cournoyer*
Université de Montréal

La création de bases de données locales à partir d'un télédéchargement suscite de plus en plus l'intérêt des spécialistes de l'information. Le texte qui suit analyse les différentes facettes de cette question.

Suite à un exposé historique du processus de création de bases de données par télédéchargement, les étapes sont expliquées et l'on présente quelques types de logiciel de télécommunications.

La deuxième partie comprend la définition et le rôle de la conversion. On met l'accent sur les aspects techniques qui composent le traitement des données télédéchargées. Plusieurs exemples illustrent ce procédé.

La troisième partie expose une brève revue historique de la loi américaine du droit d'auteur et aborde les problèmes légaux liés aux différents types de télédéchargement. Les préoccupations des producteurs et des serveurs sont signalées de même que les diverses actions entreprises pour tenter de contrôler le télédéchargement.

Downloading and data base creation: communication, conversion and copyright

The downloading of data for the purposes of creating local data bases is of growing interest to information professionals. The following article examines this phenomenon.

Following a historical summary, the first part of the article describes the creation of a data base after downloading. All the steps are outlined and a number of telecommunications software are described.

The second part defines and describes the conversion. An emphasis is placed on the technical aspects involved in the processing of the downloaded data. Several examples illustrate this process.

The third part briefly summarises the American Copyright law and the legal problems arising from different types of downloading. The concerns of the producers and vendors are outlined as well as the means taken to control downloading.

Teledescarga y creación de bases de datos: comunicación, conversión y derechos de autor

La creación de bases de datos locales a partir de una teledescarga suscita cada vez m's el interés de los especialitas de la información. El texto siguiente analiza las diferentes facetas de este tema.

Después de un breve resumen histórico, la primera parte de este artículo describe el proceso de creación de bases de datos por teledescarga. Se explican todas las etapas de esta technolog!a y se presentan algunos tipos de sistemas de programación para las telecomunicaciones.

La segunda parte incluye la definición y el papel de la conversión. Se pone el énfasis sobre los aspectos técnicos que componen el tratamiento de datos teledescargados. Muchos ejemplos ilustran este proceso.

La tercera parte presenta una breve revisión histórica de la ley americana en cuanto a los derechos de autor y trata de los problemas legales referentes a varios tipos de teledescarga. También hace resaltar las preocupaciones de los productores y de los vendedores así como las diversas acciones elaboradas para intentar el control de la teledescarga.

* Yvette Henry-Rousseau est professeur adjoint et Fethy Mili et Sylvie Cournoyer diplômés de l'École de bibliothéconomie et des sciences de l'information de l'Université de Montréal.

Les bases de données locales peuvent être définies comme des bases de données que l'on développe pour répondre à des besoins spécifiques d'un centre de documentation spécialisé, d'une équipe de recherche ou d'un individu. Les informations qui y sont contenues peuvent être de plusieurs types: bibliographiques (les plus courantes), numériques ou textuelles (texte intégral). Elles se caractérisent par un nombre limité d'usagers, un nombre réduit de documents (quelques milliers au maximum) et par leur accessibilité sur micro-ordinateur à l'aide de logiciels (documentaires ou non) disponibles sur le marché. On les utilise pour effectuer des recherches automatisées d'information ou pour éditer des bibliographies ou des rapports techniques.

Un bref aperçu historique permet de mieux comprendre l'évolution du processus de création de ces bases de données pour usage local. Il y a quelques années, les centres de documentation spécialisés développaient, pour leurs usagers, des outils bibliographiques tels les listes des nouveaux documents (monographies, articles de périodiques, etc.) qui pouvaient déjà être considérés comme des minibases de données. Certaines de ces listes pouvaient être personnalisées et répondre à des profils d'intérêt particuliers. La plupart du temps, cette activité était manuelle et exigeait que les bibliothécaires y consacrent une part importante de leur temps, ne serait-ce que pour la préparation matérielle de ces listes. Faute de temps, ce service n'était donc pas toujours aussi complet que l'aurait voulu le bibliothécaire. C'est une situation que l'on retrouve encore assez souvent dans les centres non informatisés.

Ensuite, avec la venue de services comme CAN/SDI, de l'Institut canadien d'information scientifique et technique, qui offrait des profils automatisés à partir de grandes bases de données, les usagers de ces centres de documentation ont pu recevoir régulièrement des références bibliographiques de documents qui n'appartenaient pas à la collection du centre mais qui étaient très importants pour leurs activités de recherche ou de développement. Petit à petit, les centres documentaires se constituaient ainsi des bases de données bibliographiques sur papier qu'il leur était malheureusement difficile de consulter.

Puis, il est devenu possible d'effectuer des recherches bibliographiques rétrospectives en ligne en interrogeant, à partir d'un terminal simple (écran, imprimante et modem), les grands serveurs bibliographiques (DIALOG, SDC, NLM, BRS, QUESTEL, etc). Là encore, les résultats étaient obtenus sur support papier et par conséquent restaient difficiles à réutiliser. Lorsque ces serveurs ont permis à leurs usagers de créer leurs propres profils, ils ont ouvert la porte à la création de véritables bases de données locales. En effet, les usagers pouvaient d'abord faire une recherche rétrospective sur leur sujet, puis la compléter au fur et à mesure des mises à jour des grandes bases de données commerciales qui couvrent, en général, la littérature internationale dans les domaines d'intérêt des usagers d'un centre de documentation plus exhaustivement que celui-ci ne pourrait le faire. Cela a permis de consacrer plus de temps au traitement et à la diffusion de la documentation interne, locale ou très spécifique.

Parallèlement, certains centres de documentation, surtout ceux qui servaient des entreprises informatisées, ont développé, en utilisant les ressources d'un ordinateur central, des bases de données internes dans lesquelles il était possible d'effectuer des recherches rétrospectives.

De son côté, l'informatique subissait la révolution de la micro-informatique et offrait des possibilités d'automatisation, à coût assez raisonnable, aux petites unités documentaires, permettant à celles-ci d'utiliser les micro-ordinateurs en remplacement des terminaux traditionnels et de les utiliser comme ordinateurs locaux pouvant réaliser des tâches bibliothéconomiques spécifiques, notamment la création de bases de données. Ce dernier aspect s'est développé surtout lorsque les disques rigides ont pu être couplés aux micro-ordinateurs. La machine à traitement de textes, puis le logiciel de traitement de textes ont favorisé l'éclosion de besoins de recherche automatisée d'information dans les textes intégraux. C'est ainsi que certains professionnels comme les avocats et les notaires sont en train de créer des bases de données locales de texte intégral. On peut pratiquement penser maintenant que la création et l'interrogation de bases de données locales vont devenir un problème d'actualité pour tous les usagers de la bureautique.

On avait, d'une part, des besoins et des habitudes de référence automatisée et, d'autre part, des possibilités très grandes offertes par la micro-informatique en termes de télécommunication et de traitement informatisé. C'est donc tout naturellement que l'on a pensé à réunir, sur les mêmes équipements informatiques, les données traitées localement et celles reçues par le micro-ordinateur «terminal» afin de constituer une seule base de données réutilisable localement. C'est à partir de 1982 que l'on a vu apparaître ce nouveau phénomène appelé télédéchargement ou «downloading». Les années 1984 et 1985 ont été des années de pointe pour le développement des logiciels de télécommunication, assez faciles à manipuler, qui permettent un télédéchargement des données. Quant à la

réutilisation des données, il semble que l'on soit en pleine période de création et de commercialisation de systèmes universels (logiciels qui traitent des données provenant de bases ou de serveurs différents).

Nous allons d'abord voir, dans une première partie, le rôle du téléchargement et du télédéchargement dans le processus de création de bases de données et les problèmes particuliers de télécommunication; dans un deuxième temps, nous aborderons d'une manière plus spécifique les aspects techniques de la conversion de données télédéchargées; enfin, dans une troisième partie, nous analyserons les problèmes de droit d'auteur soulevés par l'utilisation des données ainsi obtenues.

Création de bases de données par télédéchargement

Ce processus d'acquisition de données était peu connu en bibliothéconomie jusqu'à l'arrivée des micro-ordinateurs. Par contre, le transfert de données d'un ordinateur vers un autre ordinateur pour télétraitement est pratiqué depuis longtemps en informatique. C'est la première application importante des réseaux informatiques.

Avant de traiter en détail le processus du télédéchargement, nous allons voir son rôle dans la création de bases de données. Rappelons d'abord brièvement les principales étapes du processus de création de bases de données, quel que soit l'ordinateur utilisé:

Phase I: définition de la structure de la base de données, c'est-à-dire choix des éléments d'informations à prendre en compte pour chaque type de document, regroupement de ces informations en champs (ou zones) et sous-champs, puis mise en ordre de ces champs par formulaires ou bordereaux.

Phase II: saisie des données par appel de formulaires vierges et des champs correspondants à partir des informations relatives au document analysé.

Phase III: validation des informations entrées sous forme de contrôle sémantique (vocabulaire d'indexation contrôlé), de contrôle syntaxique (vérification des formes d'écriture de certaines zones), etc.

Phase IV: chargement de la base pour interrogation, c'est-à-dire création des différents fichiers inversés qui permettront la recherche en direct.

Le télédéchargement remplace, en partie, la phase II (saisie des informations à intégrer à la base). C'est bien sûr une économie de temps et d'argent au niveau strictement informatique. Les coûts de saisie se situent généralement entre deux et quatre dollars par mille caractères. Toutefois, les gains les plus importants se trouvent au niveau des traitements spécifiques et analytiques.

Une fois les données stockées sur disque, on a souvent besoin de les manipuler avant de les intégrer dans la base locale. Le degré de traitement qu'on peut leur faire subir dépend, bien sûr, de l'utilisation que l'on veut en faire. On identifie plusieurs types de réutilisation des données télédéchargées:

— les références obtenues sont enregistrées dans une mémoire de masse (disquette ou disque rigide) au cours de la session d'interrogation, puis, lorsque le micro-ordinateur n'est plus relié au serveur, elles sont imprimées localement. Cette approche réduit les coûts des sessions d'interrogation tout en permettant d'obtenir directement toutes les références retrouvées. Le micro-ordinateur peut, en effet, être équipé d'un modem à plus grande vitesse (1 200 à 2 400 bauds) que ceux que l'on retrouve généralement couplés aux terminaux esclaves (300 bauds). Certains logiciels d'interface permettent même d'éliminer tout dialogue avec l'ordinateur afin de ne conserver et d'éditer que les références bibliographiques;

— les références obtenues sont enregistrées au fur et à mesure du déroulement de la recherche en ligne. Cette étape étant renouvelée sur autant de bases de données et de serveurs que l'on veut, on peut vouloir les fusionner afin d'éditer une bibliographie complète sur un sujet donné. Certains logiciels d'édition de bibliographies permettent de réécrire toutes les notices dans un même format et d'éliminer les doubles. D'autres vont même jusqu'à offrir la constitution d'index pour compléter l'édition de la bibliographie;

— les références obtenues sont enregistrées à chaque session d'interrogation puis traitées localement afin d'être fusionnées à une base de données interne. Cette base de données est interrogeable à partir du micro-ordinateur ou d'un ordinateur central. C'est cet aspect que nous analyserons plus en profondeur.

Le processus de télédéchargement

Le télédéchargement est le transfert de données d'un ordinateur vers un autre. Dans le cas qui nous intéresse, on peut identifier deux étapes principales: 1) l'interrogation d'un serveur de bases de données et le stockage des résultats de la recherche; et, 2) les traitements que l'on fait subir aux informations obtenues avant de les entrer dans le système local.

En micro-informatique, on a donné un sens au transfert de données selon que l'envoi provient

d'un micro-ordinateur «terminal» ou d'un ordinateur central (mainframe computer). Ainsi, on parlera de *télédéchargement* (downloading) si les informations proviennent de l'ordinateur central et sont reçues par le micro-ordinateur et de *téléchargement* (uploading) dans le sens contraire. C'est pourquoi, lorsque l'on interroge des serveurs de bases de données dans le but de conserver (provisoirement ou à long terme) certains résultats de recherche, on utilise les deux sens de la communication. Ce point est important car il détermine le choix du logiciel de télécommunication dont il faudra munir le micro-ordinateur pour qu'il fonctionne tel qu'on le désire.

Logiciels de télécommunication

Un logiciel ordinaire transformera le micro-ordinateur en terminal non intelligent traditionnel, c'est-à-dire que le branchement et le déroulement de la recherche s'effectueront à partir du clavier du micro-ordinateur et que les résultats s'éditeront sur l'imprimante. Le micro-ordinateur se comporte dans ce cas comme un simple terminal-écran avec imprimante. Pour pouvoir télédécharger, il est nécessaire d'avoir un logiciel qui permette d'écrire les informations reçues, au choix, sur imprimante ou sur disque (disquette ou disque rigide). Généralement avec de tels logiciels, on peut aussi se brancher automatiquement et envoyer des fichiers créés localement à l'ordinateur principal.

Deux familles de logiciels offrent ces fonctions: les logiciels qui envoient les fichiers en bloc et ceux qui peuvent les envoyer ligne par ligne. Ces derniers sont les plus utiles en téléréférence car ils n'obligent pas l'usager à prévoir des «pauses» artificielles pour combler les temps d'attente dus à la charge du réseau informatique. De plus, en affichant la prochaine ligne à transmettre, ils laissent le choix à l'usager de l'envoyer ou de l'annuler, facilitant ainsi l'insertion de commandes non prévues au départ et permettant de réagir en fonction des résultats intermédiaires obtenus.

Branchement automatique

Grâce aux logiciels de télécommunication, il est possible d'enregistrer les procédures de connexion à plusieurs ordinateurs «serveurs». Cela se fait en général dans des fichiers distincts dans lesquels on inscrit ligne à ligne, pour chaque serveur: le numéro de téléphone à composer; les codes de reconnaissance d'équipement terminal par le réseau téléinformatique (par exemple: «.» suivi du retour de chariot dans le cas de DATAPAC) si nécessaire; l'adresse de l'ordinateur serveur sur le réseau téléinformatique et, enfin, le mot de passe habituel.

Certains de ces logiciels peuvent recomposer automatiquement le numéro de téléphone jusqu'à l'obtention de la communication ou jusqu'à un nombre maximum d'essais.

Interrogation en ligne: téléchargement

Quand l'ordinateur central a reconnu le mot de passe, il est possible de lancer l'interrogation, soit en procédant de manière traditionnelle, soit en envoyant ligne par ligne le fichier contenant des stratégies de recherche établies à priori.

La technique visant à préparer à l'avance ses recherches et à stocker les différentes stratégies corresponsantes était pratiquée, il y a quelques années, dans certains pays d'Europe qui voulaient interroger les serveurs américains, dans le temps où les connexions se faisaient via le réseau téléphonique ordinaire et où une différence de quelques minutes de branchement avait une grande influence sur le coût de la recherche. Les stratégies de plusieurs recherches étaient enregistrées sur ruban perforé (!) que les terminaux «télétypes» lisaient et envoyaient à l'ordinateur. On pouvait reprocher à cette approche d'annuler les possibilités de dialogue, mais il est certain que les bibliothécaires préparaient leurs recherches avec beaucoup de soin, ne pouvant miser sur l'ajustement pendant la session d'interrogation. Cette pratique avait presque disparu avec la venue des réseaux de téléinformatique.

On pourrait être tenté de dire qu'avec les micro-ordinateurs, on rétrograde et retourne vers d'anciennes pratiques annulant toute l'évolution du dialogue dans la téléréférence. Il n'en est rien, heureusement, à condition de bien choisir son logiciel de communication. Là encore, les Européens ont fait figure de pionniers comme le souligne James A. Benson dans l'introduction du numéro spécial de *Library HI TECH*[1] consacré au téléchargement:

> European librarians, with their higher telecommunication costs, were quick to exploit uploading of search statements to save costs[2].

Les fichiers à transmettre peuvent être créés à l'aide de l'éditeur du micro-ordinateur ou de tout autre logiciel permettant la saisie locale comme les logiciels de traitement de texte. Une

1 «Downloading/uploading online databases & catalogs», *Library HI TECH Special Studies*, no. 1 (February 1985).

2. James A. Benson, «Introduction», *Library HI TECH Special Studies*, no. 1 (February 1985), VII.

seule contrainte: les fichiers doivent être codés en ASCII (code international pour les télécommunications). Certains logiciels, surtout ceux parmi les plus anciens, ne permettent pas la création de fichiers ASCII.

Un des avantages de l'interrogation par téléchargement est la sauvegarde locale des questions pour la diffusion sélective de l'information (DSI). Cela évite de payer des frais de stockage sur les disques du centre serveur. Dans le cas d'insertion des données recueillies dans une base de données locale, la recherche courante est au moins tout aussi importante que la recherche rétrospective.

Un autre avantage du téléchargement est la réduction des temps de connexion avec les différents fichiers de l'ordinateur serveur. Ainsi, le fait d'avoir à préparer d'une manière précise (ligne à ligne) sa stratégie de recherche incite le bibliothécaire à déplacer une partie de ses demandes d'assistance en conversationnel (consultation de lexiques, de thésaurus, etc.) et à se renseigner avant le branchement.

Le téléchargement est aussi utilisé pour la mise à jour des bases de données sur l'ordinateur central à partir d'enregistrements faits localement. Cette pratique est courante pour mettre à jour certains types de bases de données qui nécessitent des mises à jour fréquentes comme les bases de données de presse. Par contre, on ne peut faire un tel téléchargement sans qu'il y ait un contrat entre producteur et serveur. Les informations reçues par l'ordinateur central doivent être disposées selon le format d'entrée des données du serveur. Il faut donc souvent faire appel à un programme de mise en forme avant l'envoi du fichier local.

Interrogation en ligne: télédéchargement

Au cours de la session d'interrogation en direct il est possible soit d'enregistrer au complet les commandes des résultats obtenus, soit d'enregistrer uniquement les références qui semblent pertinentes.

On peut utiliser le premier cas pour aider à la formulation quasi définitive de stratégies de recherche, cette étape pouvant être vue comme une interrogation préparatoire. Ce type de télédéchargement global peut être aussi considéré comme un support à la formation en téléréférence. Il suffit d'éditer la session complète, d'analyser les diverses étapes, d'apporter les corrections nécessaires dans le fichier de réception, de se rebrancher avec le serveur et d'envoyer ligne à ligne le fichier corrigé.

Le deuxième cas est le plus souvent pratiqué pour la création de bases de données locales. On est, en général, sûr de soi car ce n'est pas le premier essai que l'on fait. La seule pratique recommandable toutefois, est d'enregistrer l'historique de la recherche avant les résultats correspondants. Il est aussi très fortement conseillé d'enregistrer, dans des fichiers distincts, les résultats provenant de bases de données différentes. En effet, les informations reçues ont une structure qui varie d'une base de données à l'autre et lorsque l'on voudra les intégrer, soit pour éditer une bibliographie, soit pour les fusionner à la base locale, les traitements à leur faire subir varieront selon leur provenance.

C'est ainsi que s'arrête le processus dit de télédéchargement et pourtant les données reçues sont loin d'être immédiatement «comestibles». En fait, on peut dire que la partie «télé» est terminée et que commencent les séries de manipulations locales appelées reformatage ou conversion.

Toutefois avant de terminer la présente section, il serait intéressant de présenter quelques logiciels disponibles sur le marché qui permettent de faire du télédéchargement et qui sont loin d'offrir les mêmes possibilités.

Principaux types de logiciels pour le télédéchargement

Tout d'abord, notons les simples logiciels de télécommunication qui permettent de se brancher automatiquement à un nombre quasi illimité de serveurs (la seule contrainte pourrait être d'ordre technique), d'envoyer des fichiers globalement ou ligne à ligne et de recevoir des informations soit sur imprimante, soit sur mémoire de masse. Ils nécessitent l'achat ou le développement de logiciels complémentaires pour la conversion et la gestion de la base de données locale.

Les autres logiciels sont de type intégré, c'est-à-dire que l'on peut télédécharger des données à partir de serveurs, les fusionner à la base locale et faire des interrogations localement. La principale différence réside dans le choix du serveur ou des bases de données interrogeables. On a donc: 1) les systèmes dédiés à une seule base (Microsearch pour Eric); 2) les systèmes qui permettent d'interroger un ensemble défini de serveurs et de bases de données (Sci-Mate, Micro-Questel) et, 3) les systèmes dits «universels» qui offrent les larges possibilités des logiciels de télécommunication ainsi que des modules de reformatage et de manipulations locales (PBS-Bibliolink, Edibase). Ils sont en général un peu plus complexes à utiliser dans le sens où ils demandent à l'usager d'analyser la structure des informations reçues et de donner les règles de conversion pour leur fusion avec la base locale. Ce qui peut donc sembler une contrainte au départ est en fait un avantage car, de cette manière,

il est possible d'obtenir le produit que l'on veut dans le format que l'on veut. Un signe de la tendance vers l'universalité est que PBS vient tout juste de passer de la deuxième classe à la troisième et il semblerait que ISI soit sur la même voie.

La conversion

Comme nous venons de le voir, le télédéchargement, considéré dans son sens le plus large, n'est pas simplement l'acheminement de données d'un ordinateur central vers un poste terminal. En fait, il existe d'autres tâches complémentaires à cette opération et la conversion peut être vue comme une conséquence technique du télédéchargement. Cette conversion, qui peut passer inaperçue dans les systèmes dédiés, est absolument nécessaire pour la réutilisation locale des données reçues à partir d'un serveur.

Nous détaillerons maintenant le processus de traitement local des données récupérées. Le développement rapide des bases de données locales intensifie le rôle de cette fonction lors de la création et de la mise à jour de telles bases. Il est donc devenu nécessaire de comprendre les diverses étapes du processus et les tâches qui incombent aux spécialistes de l'information à ce niveau.

La littérature spécialisée a souvent attiré l'attention sur les retombées juridiques du télédéchargement; mais rares sont les articles qui se sont intéressés au télédéchargement comme phénomène global et surtout à ses aspects techniques. En effet, l'opération ne finit pas avec le stockage de données reçues. Les premiers logiciels qui permettaient de télédécharger étaient du type dédié ou limité et tout ce qui pouvait passer pour de la «cuisine interne» était transparent à l'utilisateur. La tendance étant de se diriger vers les systèmes «universels», qui offrent beaucoup de souplesse (un avantage pour les uns, un inconvénient pour les autres), oblige l'usager à mieux connaître la structure informatique du produit reçu et de son propre produit local.

Conversion: aspects généraux

Qu'est-ce que la conversion? Avant d'en donner une définition, il est important d'ouvrir une parenthèse terminologique. Les termes «conversion» et «reformatage» sont utilisés couramment pour désigner le même processus dans les textes spécialisés et nous suivrons cette tendance. Ceci dit, il est difficile de définir cette composante du télédéchargement surtout que l'opération elle-même se déroule d'une manière automatique. En effet le reformatage est une tâche technique exécutée par l'ordinateur en utilisant un programme et des règles de conversion. Néanmoins, avant l'exécution du programme, le gestionnaire doit procéder à quelques préparatifs (préparation des paramètres de conversion) pour donner à l'information convertie un format qui réponde aux exigences locales. C'est pourquoi il est utile de définir la conversion en fonction des différents préparatifs, le niveau automatique lui-même étant défini comme «the manipulation of the downloaded records»[3].

Réorganiser les données est une tâche nécessaire si on veut éliminer les éléments superflus en vue d'une présentation claire de ces données et d'une harmonisation du produit «brut» avec celui de la base locale. Ainsi la conversion assure, pour le gestionnaire des bases locales, une continuité dans la présentation du produit, car les formats de diffusion de serveurs ne répondent généralement pas aux exigences de la base locale. La figure 1 démontre à quel point les formats d'affichage tels que structurés par les serveurs peuvent être complexes à déchiffrer pour des usagers peu expérimentés.

Selon James A. Benson, la conversion des données est indispensable:

> When we download data, we most commonly capture them as plain text. The beginning and end of records and fields are not marked with special codes. This means that before we can insert the data into an application program such as a database management system or an electronic spreadsheet, the data must be structured, formatted, and translated into required code so that they can be used by such software[4].

Les principaux objectifs de la conversion sont: 1) l'harmonisation des formats d'affichage vu les différences qui existent entre les formats des données transmises par les serveurs; 2) le nettoyage du fichier qui consiste à libérer le terminal de certaines chaînes de caractères ou de messages «parasites» qui perturbent la disposition de l'information; et, 3), l'adaptation des données aux besoins des usagers et à la structure de la base locale. Cette opération peut comprendre la réorganisation des données

3. Victor Rosenberg and Peggy Benson, «The professional bibliographic system and Biblio-Link: information management for downloaded records», *Library HI TECH Special Studies*, no. 1 (February 1985), 52.

4. James A. Benson, «Terminology of downloading / uploading», *Library HI TECH Special Studies*, no. 1 (February 1985), 5.

```
 8/5/6
1180547   ELW
   Tranomat Interface Re-Formats Any Drill Program Generated.
   Anon
   Electronics Weekly  No.1289, Oct. 16, 1985, P. 43. 1 Page.
   COUNTRY OF PUBLICATION: United Kingdom   LANGUAGE: English
   DOCUMENT TYPE: Tabloid   ARTICLE TYPE: Product Announcement
   A drill program interface system that translates drilling information
from one format to another, the Tranomat from Dorniver can reformat
programs generated on a drill, remote programmer, or CAD/CAM system. The
device consists of a stand-alone computer with either serial or parallel
communications, cassette storage, a keyboard, and a liquid crystal display.
The system will accept drilling information from a variety of sources
including programs from leading drill manufacturers worldwide.
   PRODUCT NAME: Tranomat ,    , Dorniver , Special-Purpose Computers
   DESCRIPTORS:  CAD/CAM; Manufacturers; Fully Automated; Machinery;
Intelligent Devices; Format; Software; Process Control; Automation; New
Product; Special-Purpose Computers
```

UNE NOTICE TELLE QUE TELEDECHARGEE DU SERVEUR DIALOG.
ON REMARQUE LE NIVEAU D'OPACITE QUI CARACTERISE LA DISPOSITION DES DONNEES.

(changement de la disposition des étiquettes ou des noms des zones), l'ajout de nouveaux champs, l'élimination de certains champs jugés superflus, la suppression ou la traduction de chaînes délimitant le début d'un nouvel enregistrement (Exemple: «FN»de DIALOG).

Plusieurs autres changements peuvent être introduits pour assurer l'harmonisation des données télédéchargées à la structure initiale de la base locale. Selon Charles M. Goldstein et Maureen Prettyman, les principales manipulations à l'étape post-édition («post-processing») sont les suivantes:

A— Translate fields labels in all fields to a common set.
B— nclude for, and add, database and retrieval system names to all records.
C— Merge all records into one file.
D— Reorder the records into a format that is optimized for further processing.
E— Determine and add the type of publication[5].

Conversion: aspects techniques

Pour expliquer les différentes étapes de la conversion des notices bibliographiques, nous analyserons les manipulations que devront subir ces notices selon qu'elles seront incorporées à une base de données locale ou qu'elles serviront à élaborer une bibliographie.

Les références utilisées comme exemples proviennent de télédéchargements faits à partir de trois serveurs: CAN/OLE, DIALOG et QUESTEL. Nous nous sommes limités à ces trois serveurs mais une généralisation serait toujours possible. D'ailleurs, le but poursuivi ici n'est pas de dresser un bilan de tous les formats de tous les serveurs mais bien d'illustrer les principales manipulations que l'on peut avoir à effectuer au cours de la conversion.

Nous avons choisi de télédécharger des références bibliographiques de DIALOG en utilisant le nouveau format 4. Disponible depuis quelques mois seulement, ce format d'affichage

5. Charles M. Goldstein and Maureen Prettyman, «Processing downloaded citations», *Library HI TECH Special Studies*, no. 1 (February 1985), 46.

n'est pas valide pour toutes les bases de données. Il offre une présentation des informations avec des étiquettes de champs facilement identifiables par un programme de traitement local, la fin de chaque champ étant identifiée par un caractère («ù» dans notre exemple) qui sert de délimiteur de champ.

D'autres serveurs favorisent le télédéchargement comme le serveur IRS de l'Agence Spatiale Européenne avec sa commande DOWNLOAD et son FORMAT X:

> The ESA-IRS FORMAT X has been specially prepared to facilitate the DOWNLOAD of database material from the ESA-IRS mainframe onto microcomputers, word processors, or any form of user private magnetic support. The characteristic of FORMAT X is that downloaded data are received by the microcomputer under the form of a continuous string of alphanumeric characters where all the fields present in a given item are explicitly tagged. The format also contains the necessary symbols for proper recognition of text start or end as well as subfield or sub-subfield indicators[6].

Nous avons vu précédemment que le gestionnaire de la base doit effectuer un certain nombre de préparatifs avant la conversion. La tâche principale, à ce niveau, est l'analyse détaillée des informations reçues et stockées. C'est-à-dire que le gestionnaire doit décider quelles informations seront conservées pour un versement dans la base locale, vérifier quels critères le programme de conversion pourra utiliser pour cette sélection et, enfin, s'assurer que l'élément qui doit subir les changements est clairement et uniformément repérable dans chaque document du fichier télédéchargé. En premier lieu, il faut donc qu'il identifie d'une manière unique les caractères qui permettent de reconnaître le début et la fin d'un enregistrement. Pour ce faire, il peut être important de posséder un programme d'analyse de fichiers qui permette de «voir» les caractères de contrôle (fins de ligne, séquences spéciales, etc.) et les codes utilisés pour les caractères accentués lorsque disponibles.

Ces caractères ne sont pas toujours visibles car ils ne s'impriment pas, mais ils possèdent une valeur binaire que l'on peut afficher avec un tel programme d'analyse.

La conversion peut se faire de plusieurs façons selon le type de logiciel utilisé. Par exemple, il est possible de reformater à la main les données recueillies en utilisant un logiciel de traitement de texte (à la condition qu'il traite des fichiers ASCII). C'est ainsi que le souligne Kenneth R. Murr:

> The use of a word processing program to edit the results is probably the most commonly used way to manipulate the search results. The advantage is that one can use the standard program that one is familiar with to eliminate those infamous false drops that occur in every search... The main disadvantage of using a word processing program is the amount of time it takes to edit the material... Unfortunately, they are not well suited to this type of editing[7].

Il est bien entendu que ce genre de manipulations peut être très fastidieux et doit être répété à chaque télédéchargement. Par ailleurs, si l'on fait appel à un programme de reformatage, l'effort d'analyse est peut-être un peu plus complexe, mais, une fois établis les critères de conversion, l'exécution de la transformation des données se fait automatiquement et peut se répéter à chaque télédéchargement tant que les formats externes ou internes ne changent pas.

Un programme de reformatage doit permettre d'utiliser les caractères de contrôle car la plupart des noms de zones à identifier commencent en début de ligne. Il est important de pouvoir spécifier que l'on recherche une séquence: «fin de ligne, nom de zone:», et non seulement le nom de la zone qui pourrait se retrouver n'importe où dans le texte (exemple: «au:» et «traitement de l'eau:»). Les caractères de contrôle sont de plus indispensables pour la présentation des données remodelées (fins de ligne, sauts de ligne, etc.).

Un autre aspect technique important de la conversion qui est relié au codage de l'information reçue est le transcodage. Par transcodage on entend remplacement de chacun des caractères reçus par un autre ou élimination de ce caractère. Cela peut servir à traiter les caractères français de plusieurs serveurs, même s'ils utilisent des codes différents (QUESTEL, MINISIS, etc.) ou encore de traiter d'autres codes que l'ASCII (exemple: code EBCDIC utilisé par l'Office de la langue française), et bien sûr d'éliminer les caractères inutiles transmis parfois par les serveurs. Ce transcodage peut être fait soit par le logiciel de télécommunication, ce qui représente un avantage car les caractères reçus apparaissent correctement à l'écran, soit par le logiciel de reformatage à condition qu'il dispose d'une table de transcodage. Cette méthode peut être très utile pour le téléchargement d'une base de données locale sur un serveur.

6. James A. Benson and Bella H. Weinberg, «Downloading policies», *Library HI TECH Special Studies*, no. 1 (February 1985), 82.

7. Kenneth R. Murr, «After downloading; how to get the most out of your microcomputer», in *National Online Meetings 1984, Proceedings of the fifth National Online Meeting*, New York, April 10-12 1984, p. 244.

Principes de fonctionnement du logiciel utilisé

Le logiciel qui a été utilisé pour fournir les différents exemples est EDIBASE (module de télécommunication et module de conversion). Afin de mieux suivre les diverses étapes de la conversion, voyons les tâches que ce logiciel peut accomplir:

— remplacer une chaîne de caractères par une autre chaîne;
— éliminer des portions de texte;
— transcoder des caractères;
— insérer des séparateurs de documents;
— insérer un même texte à la fin de tous les documents.

Chaque logiciel a ses propres règles d'organisation de ses bases de données. Nous utiliserons dans cette étude la structure d'une base de données telle que définie par le logiciel EDIBASE:

Les zones qui délimitent chaque document dans la base sont: la zone NO (numéro d'identification du document) qui marque le début de chaque document et la zone FIN qui représente la chaîne de caractères finissant le document; Tous les noms de zones sont inscrits entre deux tirets sur une ligne; il faut se positionner au niveau de la deuxième ligne pour écrire les données (le contenu de la zone)[8].

Les informations concernant le reformatage sont enregistrées dans un fichier, créé par le gestionnaire de la base, et qui comprend les intructions de remplacement, d'élimination d'une chaîne de caractères, de transcodage, d'insertion ou d'ajout d'une chaîne de caractères. D'autres instructions peuvent être incorporées au fichier de conversion[9]. Certaines sont d'ailleurs utilisées dans les modèles de fichier de conversion (figures 3A, 3B et 3C).

Exemples de conversion

Comme on l'a déjà dit, avant de passer au versement des données dans la base, le gestionnaire doit analyser son produit brut. Les figures 2A, 2B et 2C représentent des notices telles que télédéchargées. Au cours de l'analyse, le gestion-

8. *Manuel d'utilisation du logiciel EDIBASE Version 1.0*, Montréal, INFORM II, 1985.

9. *Ibid.*

UNE NOTICE TELEDECHARGEE DE DIALOG

```
5/4/1
(FN- DIALOG AGRICOLA File 10ù)
(N- 86020979ù)
AN- <AGRICOLA Number> 85039703ù
TI- Acidic pollutants in air and precipitation at selected rural locations
    in Canadaù
AU- Barrie, L.A.ù
AU- Anlauf,K.; Wiebe, H.A.; Fellin, P.ù
JN- Desposition both wet and dry/ edited by bruce B. Hicks.ù
CL- P. 15-35. ill., maps.ù
PU- Boston : butterworth Publishers, c1984.ù
SE- Acid precipitation series ; v. 4ù
BN- 0250405695ù
CA- <DNAL> TD196.A25D45ù
HL- AGLù
LA- Englishù
NT- Paper presented at the "ACS Acid Rain Symposium" held March 28, 1982,
     in Las Vegas, Nevada.ù
NT- Includes 16 references.ù
SF- OTHER US  (NOT EXP STN, EXT, USDA, SINCE 12/76)ù
DT- Articleù
DE- acid rain; air pollutants; precipitation; rural areasù
GL- canadaù
SH- POLLUTION (W000)
```

FIGURE 2A

Le caractère ù marque la fin de chaque zone. Ce caractère change selon les imprimantes. A l'écran il prend la forme suivante: ?

UNE NOTICE TELEDECHARGEE DE CAN/OLE

```
RE - ZZ      IPE 580U66       ENG
FO - OORTA  D  BL A TO 3        1
TI - URBAN PUBLIC TRASPORTATION : A STRATEGY FOR EVALUATION .
OR - (IPE5) ORGANISATION FOR ECONOMIC  CO-OPERATION AND DEVELOPMENT. ROAD
       RESEARCH GROUP
SE - OECD ROAD RESEARCH.
NU - URBAN
DA - 1980
MO - INTERNATIONAL ORGANIZATIONS
SU - (TO) URBAN TRANSPORTATION
FO - (BKLT) BOOKLET (50-99 P.)
LA - (ENG) ENGLISH
```

FIGURE 2B

UNE NOTICE TELEDECHARGEE DE QUESTEL

```
(1/4 - (C) CB)
NO  : 090586-00649
TI  : Emulation du TD830 : version Systeme cle de IBM
CS  : CEGIR (Montrel, Quebec)
RE  : 8527773
SO  : Montreal : CEGIR; Manuel d'installation/reference (12p.); Prix sur
      demande
NT  : Pour ordinateurs IBM Systeme cle et Burroughs , CP/M-86; lien
      asynchrone ( jusqu'à 9600 bauds ), convertisseur TD1-Rs232 ou modem
DP  : 1983
LA  : FRE : Francais
PY  : QUE : Quebec
SL  : C : Concu au Quebec
AB  : Ce ***logiciel***permet aux terminaux du "Systeme cle IBM de
      communiquer avec tous les systemes centraux de la compagnie Burroughs.
      Il opere sous le systeme d'exploitation CP/M-86 et fonctionne avec
      un lien asynchrone (jusqu'a 9600 bauds), en mode direct RS232 ou en
      mode TDI avec convertisseur TDI-RS232 ; il peut egalement etre
      utilise en mode modem. Notons que le Systeme cle peut partager le
      lien avec d'autres terminaux Burroughs compatibles.
HS  : IBM Systeme cle; Burroughs
      (ACQ : CEGIR                                                    )
CH  : AC :Communication; E55
ID  : Emulateurs de terminaux; TD 830 (terminal); Communication;
      ***Telecommunication***
```

FIGURE 2C

naire de la base doit prélever les noms de zones délimitant chaque notice, relever les codes mnémoniques de chacune des zones pour effectuer les remplacements éventuels et repérer les caractères jugés inutiles, tels que les caractères qui marquent la fin de chaque zone.

La deuxième étape consiste à formuler le fichier de conversion, appelé aussi fichier de contrôle, qui comprend les éléments que le gestionnaire désire introduire ou éliminer du fichier de données télédéchargées. Les fichiers 3A, 3B et 3C révèlent le contenu possible du fichier de conversion. Dans les exemples fournis, le fichier débute toujours par une instruction FINDOC. Sous cette instruction, le gestionnaire signale au programme la chaîne de caractères identifiant le début d'un document télédéchargé. La deuxième ligne de cette instruction contient une nouvelle zone à ajouter à la dernière notice pour en marquer la fin.

La seconde instruction dans ce fichier est REMPLACER. Elle permet au gestionnaire d'inclure les différents remplacements qu'il compte introduire. Ainsi, dans les trois modèles de fichier de conversion, la chaîne de caractères «AU» est remplacée par «AUTEUR», Les codes '13'10 marquent la valeur en ASCII du retour chariot et le début d'une nouvelle ligne. Cette manière de disposer les zones est conforme au format exigé par EDIBASE.

Le fichier de conversion peut également contenir d'autres instructions. Ainsi, en utilisant ÉLIMINE, le gestionnaire mentionne les données (chaînes de caractères) qu'il désire éliminer.

Le fichier de la figure 3C comprend l'instruction COMMENTAIRES. Ceci permet au gestionnaire d'éliminer une portion d'un texte. Dans l'exemple mentionné dans ce chapitre, une zone a été éliminée. Le processus consiste à démarquer une chaîne de caractères, puis à marquer sous COMMENTAIRES les valeurs ASCII des caractères qui délimitent la partie du texte à supprimer. Dans le cas de l'exemple 3C, les deux caractères étaient deux accolades; leurs valeurs en code ASCII étaient respectivement 123 et 125.

La dernière instruction que comprend l'exemple 3C est TRANS (abréviation de transcodage). Elle permet de supprimer des caractères individuels. Dans l'exemple, le caractère à éliminer est l'astérisque «*».

Une fois le fichier de conversion constitué, le gestionnaire passe à l'exécution du reformatage. Le programme se charge d'exécuter les ordres mentionnés dans le fichier de conversion. On peut

MODELE DE CONVERSION DU LOGICIEL EDIBASE POUR DIALOG

```
FINDOC
"FN-"
"'13'10-FIN-'13'10"
REMPLACER
"AN- " PAR "-NO-'13'10     "
"CA- " PAR "-NO-'13'10     "
"AU- " PAR "-AUTEUR-'13'10     "
"TI- " PAR "-TITRE-'13'10     "
"JN- " PAR "-NOM/JOURNAL-'13'10     "
"PY- " PAR "-ANNEE/PUB-'13'10     "
"PU- " PAR "-EDITEUR-'13'10     "
"ED- " PAR "-EDITION-     "
"PY- " PAR "-DATE/PUB-'13'10     "
"LA- " PAR "-LANGUE-'13'10     "
"DT- " PAR "-TYPE/DOCUMENT-'13'10     "
"AB- " PAR "-RESUME-'13'10     "
"ID- " PAR "-IDENTIFICAT-'13'10     "
"LC- " PAR "-BIBLIO/CONG-'13'10     "
"HL- " PAR "-LOCAL/DOC-'13'10     "
"SH- " PAR "-CODE/SECTION-'13'10     "
"SC- " PAR "-CODE/SECTION-'13'10     "
"SN- " PAR "-ISSN-'13'10     "
"BN- " PAR "-ISBN-'13'10     "
"CS- " PAR "-SOURCE-'13'10     "
"CP- " PAR "-SOURCE-'13'10     "
"DE- " PAR "-DESCRIPTEURS-'13'10     "
"SH- " PAR "-DESCRIPTEYRS-'13'10     "
"SL- " PAR "-SOMMAIRE-'13'10     "
"SF- " PAR "-SOUS/FICHIER-'13'10     "
"CO- " PAR "-CODEN-'13'10     "
"FP- " PAR "-PAGINATION-'13'10     "
"SE- " PAR "-SERIE-'13'10     "
"CR- " PAR "-REFER/CITEES-'13'10     "
"RF- " PAR "-REFER/CITEES-'13'10     "
"GA- " PAR "-CODE-'13'10     "
"GS- " PAR "-VERIFI/ETIQ-'13'10     "
"JA- " PAR "-CODE/JOURNAL-'13'10     "
"JC- " PAR "-CODE/JOURNAL-'13'10     "
"CT- " PAR "-NOM/CONFER-'13'10     "
"CC- " PAR "-NOTES-'13'10     "
"CL- " PAR "-NOTES-'13'10     "
"GL- " PAR "-NOTES-'13'10     "
"NT- " PAR "-NOTE-'13'10     "
"BC- " PAR "-NOTES-'13'10     "
"DC- " PAR "-DOC/CODE-'13'10     "
ELIMINER
"FN- "
TRANS
124=0
COMMENTAIRES
123,125
```

FIGURE 3A

```
MODELE DE CONVERSION DU LOGICIEL
EDIBASE POUR CAN/OLE

FINDOC
"RE - "
"'13'10-FIN-'13'10"
REMPLACER
"'0'0'0'0'0'0" PAR "    "
"RE - " PAR "-NO-'13'10      "
"FO - " PAR "-FONDS-'13'10      "
"TI - " PAR "-TITRE-'13'10      "
"AU - " PAR "-AUTEUR-'13'10      "
"OR - " PAR "-ORGANISME-'13'10      "
"SE - " PAR "-SERIE-'13'10      "
"NU - " PAR "-NUMERO-'13'10      "
"DA - " PAR "-DATE-'13'10      "
"NO - " PAR "-NOTE-'13'10      "
"MO - " PAR "-MOT/CLE-'13'10      "
"SU - " PAR "-SUJET-'13'10      "
"FO - " PAR "-FORME-'13'10      "
"LA - " PAR "-LANGUE-'13'10      "
```

FIGURE 3B

```
MODELE DE CONVERSION DU LOGICIEL
EDIBASE POUR QUESTEL

FINDOC
"NO  : "
"'13'10-FIN-'13'10"
"'13'10"
REMPLACER
"NO  : " PAR "-NO-'13'10      "
"TI  : " PAR "-TITRE-'13'10      "
"CS  : " PAR "-COMPAGNIE-'13'10      "
"RE  : " PAR "-REFERENCE-'13'10      "
"DT  : " PAR "-TYPE/DOC-'13'10      "
"SO  : " PAR "-SOURCE-'13'10      "
"NT  : " PAR "-NAT/PROD'13'10    "
"DP  : " PAR "-DATE/PUB-'13'10      "
"LA  : " PAR "-LANGUE-'13'10      "
"PY  : " PAR "-PAYS/PUB-'13'10      "
"SL  : " PAR "-ORIGINE/SYS-'13'10    "
"AB  : " PAR "-RESUME-'13'10      "
"HS  : " PAR "-MOTS/CLES-'13'10    "
"ID  : " PAR "-IDENTIFICATEURS-'13'10    "
COMMENTAIRES
123,125
TRANS
42=0
```

FIGURE 3C

remarquer dans les figures 4A, 4B et 4C les changements apportés aux données télédéchargées.

Après la conversion, le gestionnaire doit effectuer une évaluation du produit obtenu, afin de s'assurer que les données ainsi remodelées s'intègreront harmonieusement à la base. Ainsi, s'il n'est pas satisfait du résultat, il lui suffit de réorganiser le fichier de conversion afin d'ajuster les changements. Cette opération doit être faite avec beaucoup de soin car, selon Rosenberg et Benson, «it is the manipulation of downloaded records that presents the challenge»[10].

En effet, le défi est difficilement surmonté dès la première tentative; c'est pourquoi il faut prendre le temps nécessaire pour effectuer un reformatage adéquat qui réponde parfaitement aux exigences techniques de la base déjà en place.

Le télédéchargement des données ne sert pas exclusivement à l'alimentation des bases de données locales. Dans certains cas, les données serviront à compiler des bibliographies personnalisées.

Les aspects techniques de l'utilisation du télédéchargement pour la compilation de bibliographies sont les mêmes que pour le versement des données dans une base locale. D'une manière générale, les bibliographies compilées à partir d'un télédéchargement sont destinées à une clientèle bien déterminée. C'est pourquoi, dans ce cas, le gestionnaire devrait analyser les données d'une manière approfondie pour concevoir la formule qui répond aux exigences du milieu desservi. Le scénario de la conversion demeure toutefois le même (fig. 5).

Le gestionnaire devrait cependant être plus exhaustif lors de la préparation du fichier de conversion. En effet, si le versement des données dans une base locale exige un fichier de conversion adapté à un modèle pré-établi, celui de la base, l'élaboration d'une bibliographie personnalisée est plutôt assujettie à la demande ponctuelle

10. Victor Rosenberg and Peggy Benson, «The professional bibliographic...», 52.

de la clientèle. Ceci exige du gestionnaire une grande perspicacité lors de l'élaboration du fichier de conversion. Idéalement, le fichier de conversion devrait présenter un contenu panoramique des spécificités des serveurs utilisés. Au cours de la conversion, le gestionnaire pourrait ajuster le contenu à la demande formulée.

Télédéchargement et droit d'auteur

La pratique du télédéchargement implique un problème qui demeure souvent méconnu des gens qui l'utilisent. En effet, aux États-Unis, les bases de données sont, comme plusieurs autres types de documents, protégées par la loi du droit d'auteur. Voici ce que cette loi stipule selon Mary Wolfe:

> A copyright is a legal right to prevent others from making use of an original work of authorship which has been appropriately published and marked with a copyright notice and which may be registered in the United States Copyright Office[11].

Ainsi, la loi protège le producteur d'une base de données en interdisant à quiconque de s'approprier ses données sans qu'il n'y ait eu d'entente entre les deux parties.

Toutefois, comme on peut facilement le constater, notamment pour les publications écrites, cette loi n'est pas toujours respectée. Nous pouvons très bien comparer les problèmes engendrés par le télédéchargement et la photocopie comme l'exprime B. Inkellis:

> Downloading is to machine-readable data what photocopying is to information embodied on paper; it is a technologically advanced form of copying[12].

Afin de se mettre à jour, les législateurs américains ont révisé la loi du droit d'auteur qui datait de 1909. Cette réforme a donné lieu au *Copyright Act of 1976* qui est entré en vigueur en 1978. Cette loi a nécessité quelque vingt ans de travail et plusieurs comités ont siégé pour parvenir à régler les problèmes causés par les

11. Mary Wolfe, «Databases and legal protection», *Online*, vol. 6, no. 3 (May 1982), 52.

12. B. Inkellis, «Legal issues of downloading online search results», in *Online '82 Conference Proceedings*, Weston, C.T. Online Inc., 1982, p. 91.

UNE NOTICE DE DIALOG REFORMATEE

```
ND-
    <AGRICOLA Number> B5039703
-TITRE-
    Acidic pollutants in air and precipitation at selected rural locations
    in Canada
-AUTEUR-
    Barrie, L.A.
-AUTEUR-
    Anlauf, K.; Wiebe, H.A.; Fellin, P.
-NOM/JOURNAL-
    Deposition both wet and dry / edited by Bruce B. Hicks.
-NOTES-
    p. 15-35. ill., maps.
-EDITEUR-
    Boston : Butterworth Publishers, c1984
-SERIE-
    Acid precipitation series ; v. 4
-ISBN-
    0250405695
-NO-
    <DNAL> TD196. A25D45
-LOCAL/DOC-
     AGL
-LANGUE-
     English
-NOTE-
     Paper presented at the "ACS Acid Rain Symposium" held March 28, 1982,
      in Las Vegas, Nevada.
-NOTE-
     Includes 16 references.
-SOUS/FICHIER-
     OTHER US (NOT EXP SIN, EXT, USDA; SINCE 12/76)
-TYPE/DOCUMENT-
     Article
-DESCRIPTEURS-
     acid rain; air pollutants; precipitation; rural areas
-NOTES-
     canada
-CODE/SECTION-
     POLLUTION (WOOD)
-FIN-
```

FIGURE 4A

UNE NOTICE DE CAN/OLE REFORMATEE

```
-NO-
     ZZ     IPE  580U66     ENG
-FONDS-
     OORTA  D  BL A TO 3     1
-TITRE-
     URBAN PUBLIC TRANSPORTATION : A STRATEGY FOR EVALUATION
-ORGANISME-
     (IPE5) ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT . ROAD
     RESEARCH GROUP
-SERIE-
     OECD ROAD RESEARCH .
-NUMERO-
     URBAN
-DATE-
     1980
-MOTS/CLES-
     INTENATIONAL ORGANIZATIONS
-SUJET-
     (TO) URBAN TRANSPORTATION
-FONDS-
     (BKLT) BOOKLET  (50-90 P.)
-LANGUE-
     (ENG) ENGLISH
-FIN-
```

FIGURE 4B

UNE NOTICE DE QUESTEL REFORMATEE

```
-NO-
    090586-00649
-TITRE-
    Emulation du TD830 : version Systeme cle de IBM
-COMPAGNIE-
    CEGIR (Montreal , Quebec)
-REFERENCE-
    8527773
-SOURCE-
    Montreal : CEGIR; Manuel d'installation/reference (12p. ); Prix sur
    demande
-NAT/PROD-
    Pour ordinateurs IBM Systeme cle et Burroughs , CP/M-86; lien
    asynchrone (jusqu'a 9600 bauds) , covertisseur TD1-RS232 ou modem
-DATE/PUB-
    1983
-LANGUE-
    FRE : Francais
-PAYS/PUB-
    QUE : Quebec
-ORIGIN/SYS-
    C : Concu au Quebec
-RESUME-
    Ce logiciel permet aux terminaux du "Systeme cle IBM" de communiquer
    avec tous les systemes centraux de la compagnie Burroughs. Il opere
    sous le systeme d'exploitation CP/M-86 et fontionne avec un lien
    asynchrone (jusqu'a 9600 bauds), en mode direct RS232 ou en mode TDI
    avec convertisseur TDI-RS232; il peut egalement etre utilise en mode
    modem. Noton que le "Systeme cle" peut partager le lien avec
    d'autres terminaux Burroughs compatibles.
-MOTS/CLES-
    IBM Systeme; Burroughs
-CLASSIFIC-
    AC : Communication; E55
-IDENTIFICATEURS-
    Emulateurs de terminaux; TD830 (terminal); Communication;
    Telecommunication
-FIN-
```

FIGURE 4C

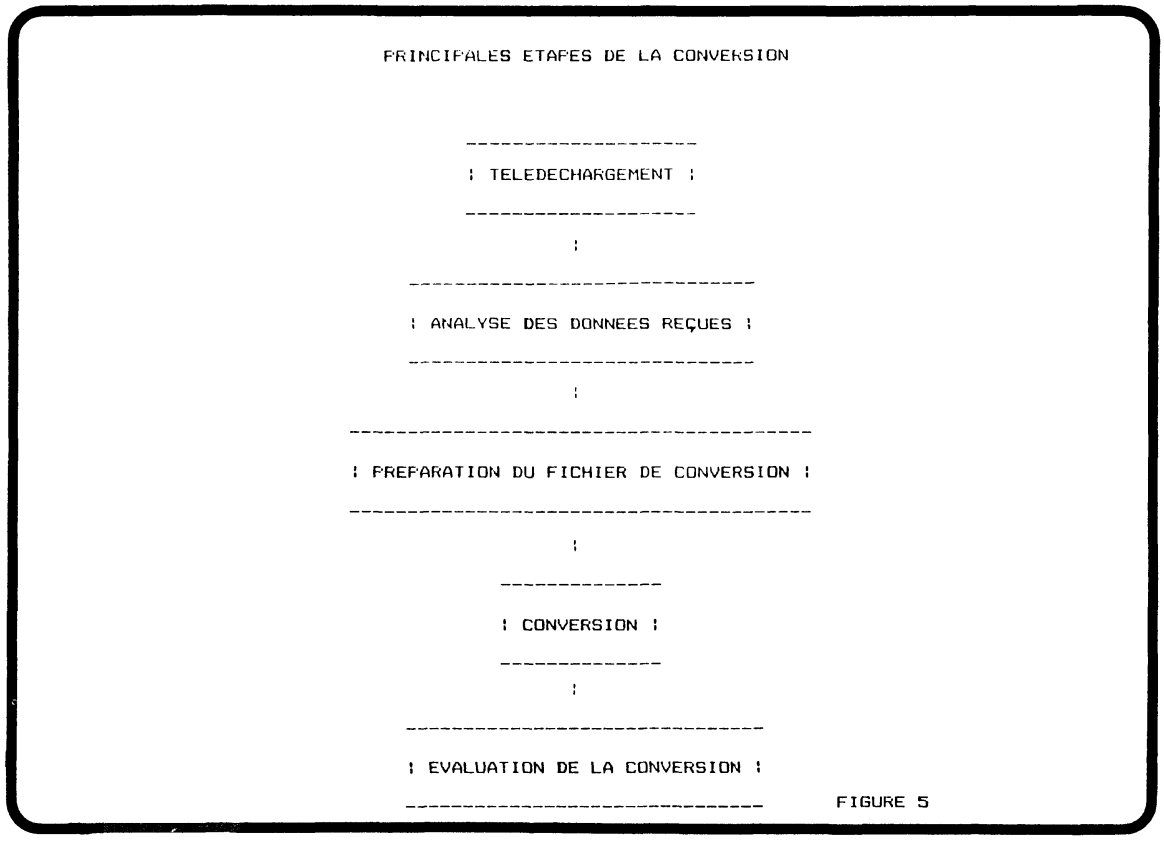

FIGURE 5

nouvelles technologies. Ainsi, la National Commission on New Technological Uses of Copyrighted Works (CONTU) s'est intéressée, entre autres, aux bases de données.

C'est principalement la section 117 qui fut discutée, celle-ci ayant été écrite pour combler les lacunes de la loi de 1909 en attendant le rapport de la commission. Mary Wolfe nous rapporte le contenu de cette section:

> Scope of exclusive rights: Use in conjonction with computers and similar information systems. Notwithstanding the provisions of sections 106 and 118, this title does not afford to the owner of copyright in a work any greater or lesser rights with respect to the use of the work in conjunction with automatic systems capable of storing, processing, retrieving, or transferring information, or in conjunction with any similar device, machine or process, than those afforded to works under the law, wheter title 17 or the common law or statutes of a State, in effect on December 31, 1977, as held applicable and construed by a court in an action brought under this title[13].

Finalement, cette section fut remodelée et ne traite maintenant que des logiciels. Pour ce qui est des bases de données, elles sont désormais régies par l'ancienne loi au même titre que les ouvrages «dérivés», les ouvrages originaux, les compilations et ce qui provient du domaine public. C'est cette situation que nous expose Jerome K. Miller:

> Copyright protection for databases falls outside the protection for computer programmes included in the new section 117. Databases receive their protection, or lack of protection, under the older provisions on public domain, compilations, derivate works, and original works[14].

De plus, cette loi comporte une clause très importante. En effet, selon l'usage qu'une personne veut en faire, l'utilisation d'un ouvrage protégé peut être considérée légale ou non. Cette clause se nomme le «fair use». Pour déterminer

13. Mary Wolfe, «Databases...», 53.

14. Jerome K. Miller, «Copyright protection for bibliographic, numeric, factual, and textual databases», *Library Trends*, vol. 32, no. 2 (Fall 1983), 200.

si l'usage est légal, il existe quatre critères spécifiques: 1) le but de l'utilisation, à savoir s'il s'agit d'un but commercial ou éducatif; 2) la nature de l'ouvrage protégé; 3) la proportion de la partie copiée par rapport à l'ouvrage original; et, 4) la portée économique de la copie sur le marché par rapport à l'ouvrage original.

Dans le contexte des bases de données, il faut donc tenir compte de l'utilisation (commerciale ou non) qui sera faite des données télédéchargées pour en arriver à porter un jugement sur la légalité de cette pratique. Il existe en effet plusieurs motifs pour faire du télédéchargement.

Divers types de télédéchargement

Comme nous l'avons vu dans la première partie, les données reçues peuvent être réutilisées de plusieurs manières: imprimer simplement les résultats des recherches sur papier; manipuler les données reçues de plusieurs bases et serveurs afin de produire des bibliographies exhaustives, éventuellement commercialisées; reformater les données reçues pour les intégrer à une base de données existante. Cette base de données peut être interrogeable par des usagers internes ou par des usagers payants.

Porter un jugement sur la légalité de ces divers types d'utilisation est très difficile. Chaque cas est particulier et mérite un examen spécial. Il est facile de penser que certaines pratiques sont tout à fait légales. Ce serait le cas du télédéchargement d'un nombre important de données sans toutefois transférer la base au complet, spécialement si l'usage prévu est non commercial ou si les données sont du domaine public. De même, le fait de télédécharger afin de modifier la présentation des données, sans qu'il y ait de conservation permanente, est tout à fait acceptable. Finalement, on peut porter le même jugement quand l'usager désire pallier à la lenteur de son imprimante en télédéchargeant les résultats de sa recherche et en les imprimant après coup, réduisant ainsi les coûts de connexion au système serveur.

Cependant, dans certaines situations, il est possible de croire qu'il y a infraction à la loi du droit d'auteur. Par exemple, si l'usage est de nature commerciale et particulièrement si les données ne sont pas d'ordre public, l'usager s'expose à plusieurs problèmes. De même, la publication d'un nouveau produit, créé à partir de données télédéchargées, peut représenter une violation de la loi. Certaines personnes s'opposent à ce jugement en disant que le produit créé est «original» du point de vue du droit d'auteur puisque les données sont réorganisées. Il est intéressant de voir ce que conseille Thomas S. Warrick à ce sujet:

> A person planning to download large amount of data from an online database for distribution to others should immediately retain a good attorney[15].

Ce qui laisse les producteurs et les usagers dans l'incertitude, c'est le fait qu'aucun jugement de cour n'ait encore été prononcé que ce soit en faveur de l'un ou de l'autre. On comprend alors la complexité des problèmes engendrés par cette pratique.

Craintes des producteurs

Suite à cette situation, les producteurs de bases de données éprouvent certaines craintes. Tout d'abord, la création de bases de données locales et la réutilisation gratuite de ces données pourraient entraîner une baisse notable de leurs profits.

De plus, la création et la vente de produits conçus à partir de leurs données, des bulletins et des bibliographies, par exemple, pourraient entrer en compétition avec leurs propres bases de données. Ce sont les deux principales craintes des producteurs. Ces craintes sont amplifiées par le fait qu'il est techniquement impossible de détecter les usagers qui font du télédéchargement.

Toutefois, certains auteurs ne sont pas d'accord avec ces arguments. Ils pensent que ces craintes sont injustifiées. Ainsi, la peur de voir apparaître des produits dérivés de leurs bases de données serait surestimée puisqu'en mettant sur le marché de tels produits, les contrevenants informeraient, par le fait même, les producteurs qu'ils ont été volés.

Un deuxième argument de taille entre en contradiction avec les craintes des producteurs. En effet, pour qu'une base de données soit «efficace», il faut qu'elle soit mise à jour régulièrement. Ainsi, le télédéchargement régulier d'une certaine quantité de données pourrait laisser soupçonner que l'usager emmagasine les données dans un but «illégal».

Il est difficile de trancher la question à savoir

15. Thomas S. Warrick, «Large databases, small computers and fast modems... An attorney looks at the legal ramifications of downloading», *Online*, vol. 8, no. 4 (July 1984), 66.

si le télédéchargement représente une menace pour la santé financière des producteurs. Ce qui est vrai, toutefois, c'est que, sauf quelques rares exceptions, les réactions face à ce phénomène ont été plutôt lentes du côté des producteurs.

Études sur le télédéchargement

Plusieurs études ont porté sur le problème du télédéchargement et l'une des plus importantes est sans contredit celle menée en 1983 et 1984 par l'équipe de John Cuadra , Cuadra multi-client study: Downloading Online Databases; Policy and Pricing Strategies. John Cuadra a réagi très rapidement face aux nouveaux problèmes du télédéchargement et invité à prendre la parole au congrès d'Atlanta, en novembre 1982, il déclarait:

> ...urged all database publishers to do some rapid policymaking about the downloading of data for storage on local microcomputers[16].

L'étude de Cuadra s'intéresse à l'opinion de toutes les parties, autant les producteurs que les clients. Parmi les 900 usagers de bases de données provenant de douze services aux Etats-Unis et au Canada, 625 ont répondu à son appel. De ce nombre, 37% ont admis pratiquer le télédéchargement pour différentes raisons, que ce soit pour une impression locale différée ou la conservation permanente pour la création d'une base de données et, parmi eux, 77% ont affirmé conserver les données temporairement, 41% voulaient reformater les données, 32% avaient l'intention de réutiliser partiellement les données et 19% ont admis vouloir conserver les données pour les réutiliser plus tard.

Mentionnons une autre étude sur le sujet, celle menée par l'Information Industry Association (IIA) auprès de serveurs et producteurs de bases de données exclusivement. Sur les 300 questionnaires distribués, les enquêteurs ont reçu 57 réponses valides qui ont servi a recueillir des informations suivantes:

— Plus de 50% des répondants ont indiqué que leur organisation avait reçu, au cours des douze derniers mois, des demandes d'autorisation de télédéchargement et de réutilisation des enregistrements;

— 25% des répondants n'avaient reçu aucune demande;

— 22% des répondants avaient reçu des demandes pour l'une ou l'autre des possibilités;

— Près de 75% des requérants voulaient utiliser l'information au niveau local plutôt qu'à des fins commerciales;

— 54% des répondants croyaient que le télédéchargement pouvait avoir un effet positif sur les revenus, 21% croyaient que l'effet serait négatif et 25% pensaient qu'il n'y aurait aucun effet sur les revenus.

Suite à ces études, les producteurs et les serveurs ont commencé à réagir et plusieurs moyens ont été mis en oeuvre pour tenter de contrôler ce problème.

Actions entreprises par les producteurs et les serveurs

Progressivement, les serveurs ont établi des contrats et des politiques spéciales pour le télédéchargement. Les ententes varient beaucoup d'une base de données à une autre comme l'indiquent les exemples suivants.

A — PSYCINFO. En 1983, le producteur de la base PSYCINFO décidait d'introduire un nouveau tarif pour l'affichage de citations dans un format complet. Il justifiait cette position en prétendant que l'usager qui affiche les citations dans le format complet procède au télédéchargement de celles-ci. Les impressions en différé coûteraient moins cher que les affichages en-ligne dans un format complet.

Cette politique pénalise les usagers qui ne font pas de télédéchargement mais qui utilisent un terminal de 1 200 bauds pour imprimer les résultats en-ligne plutôt que de les commander en différé.

B — BRS. Du côté de BRS, il a été convenu qu'à partir du milieu de 1983, un coût supplémentaire serait requis pour faire du télédéchargement. Cette politique peut difficilement être efficace car les serveurs n'ont aucun contrôle pour détecter les usagers qui utilisent cette pratique.

C — CHEMICAL ABSTRACTS SERVICE. Chez Chemical Abstracts Service, la nouvelle politique pour le télédéchargement est plus élaborée. Ainsi, les usagers des grandes institutions paieront une cotisation annuelle variant entre 4 000,00$ et 8 000,00$ pour copier jusqu'à 50 000 notices, sans restriction au niveau de l'utilisation locale. Les usagers modestes qui télédéchargent occasionnellement de petites quantités d'informations ne seront pas touchés par cette cotisation.

16. «Cuadra calls for ground-rules on downloading», *Online Review*, vol. 7, no. 1 (January 1983), 8.

Certains producteurs ont décidé de contourner le problème du télédéchargement en offrant aux usagers un autre type de service. Ce nouveau produit s'apparente au service de diffusion sélective de l'information. Il s'agit, en effet, de fournir aux usagers des «sous-bases» de données enregistrées sur disquettes et pouvant être consultées et mises à jour régulièrement par le producteur.

C'est le cas de BIOSIS qui offre le BIOSIS Information Transfer System (BITS). Les usagers peuvent choisir parmi plusieurs profils de recherche dterminés, modeler leur propre profil, ou payer BIOSIS pour les aider à en construire un. Chaque mois, un fichier mis à jour est produit selon le profil et les nouvelles citations sont envoyées à l'usager. Le tarif d'abonnement dépend du nombre de citations reçues par année. Les prix varient entre 100,00$ pour un maximum de 500 références par année et 3 200,00$ pour un maximum de 10 000 notices avec des résumés. Les abonnés peuvent choisir leur propre logiciel de recherche d'information pour interroger ou acheter le logiciel Biosuperfile de BIOSIS pour 100,00$.

Pour sa part, le producteur European Space Agency's Information Retrieval System (ESA-IRS) a décidé d'introduire une nouvelle commande «DOWNLOAD», spécifiquement pour le télédéchargement. Celle-ci fut implantée à la fin de l'année 1984 et les usagers sont maintenant encouragés à l'utiliser (légalement) plutôt que de violer la loi du droit d'auteur. Cette solution a pour avantage de faire payer uniquement les usagers qui utilisent cette technique mais pose toujours le problème de la «confiance» et n'empêche pas l'usager qui veut télédécharger des données, sans frais supplémentaires, de le faire.

Plusieurs stratégies de tarification ont été suggérées par différents auteurs, notamment Arnold A. J. Jansen. Il propose plusieurs méthodes dont celles-ci:

— Tarifs relativement élevés pour l'impression en ligne et des tarifs spéciaux pour les usagers qui acceptent de signer des accords pour pratiquer le télédéchargement. Cette situation encouragerait les usagers à légaliser leurs activités.

— Tarif de connexion bas avec des frais d'impression en ligne et en différé élevés. Cette situation permet à l'usager de déterminer le prix qu'il veut payer pour une recherche en imprimant plus ou moins de références, selon ses besoins d'information. Le producteur est payé selon le nombre de notices fournies.

— Tarif basé sur l'usage actuel du matériel, par exemple, selon le nombre de caractères imprimés. Même si le client payait en toute justice pour l'information reçue, on peut considérer cette situation comme une vente d'information à la livre. Comme quoi la quantité devient le critère principal de la valeur.

— Frais d'abonnement pour usage illimité. Cette solution offre au producteur l'avantage de disposer d'un capital à l'avance, tout comme dans le cas des ouvrages imprimés. Cette forme de tarification avantage les usagers assidus mais serait dispendieuse pour les usagers occasionnels.

Ces diverses actions entreprises par les serveurs et les producteurs devraient leur permettre d'exercer un contrôle de plus en plus serré du télédéchargement. Il n'en demeure pas moins qu'il s'agit d'un phénomène qui leur causera encore des ennuis et qui continuera à alimenter les discussions.

Conclusion

La pratique du télédéchargement est devenue chose courante pour les bibliothécaires de référence qui utilisent un micro-ordinateur comme terminal et la tendance chez les serveurs est d'en favoriser l'usage. Que ce soit pour éditer des bibliographies ou alimenter une base de données locale, le bibliothécaire devra maîtriser de plus en plus la technique de conversion et, par conséquent, l'information documentaire.

Une nouvelle approche utilisée par les producteurs de bases de données commerciales est de fournir sur disque compact[17] leurs bases de données rétrospectives accompagnées d'un logiciel d'interrogation. C'est ce que l'on appelle les bases de données CD-ROM. Le producteur envoie ensuite à ses clients, à chaque mise à jour, une copie complète de la base fusionnée. C'est une alternative intéressante au télédéchargement car, d'une part, l'usager pourra acquérir de cette manière plusieurs bases de données en toute légalité. Si les logiciels fournis par les producteurs diffèrent passablement dans leur forme, il est facile d'imaginer que l'industrie du logiciel travaillera dans le sens de l'uniformisation pour offrir sur le marché des logiciels universels. Ces nouveaux systèmes pourront d'ailleurs être vendus par les producteurs.

17 Technologie des disques optiques numériques où les informations sont lisibles par rayon laser.